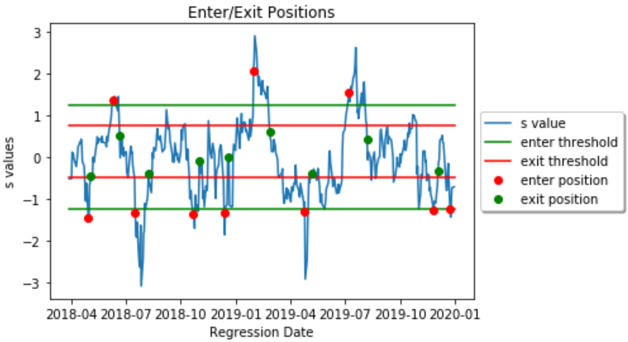
MS&E 448 Group 3: Statistical Arbitrage Strategy

Jonathan Tuck   Raphael Abbou   Vin Sachidananda

June 2020

Idea

Large positive (negative) values in a mean reverting process mean that our basket of stocks is likely to drop and produce negative (positive) returns, and we want to short (long) the basket.



Overview

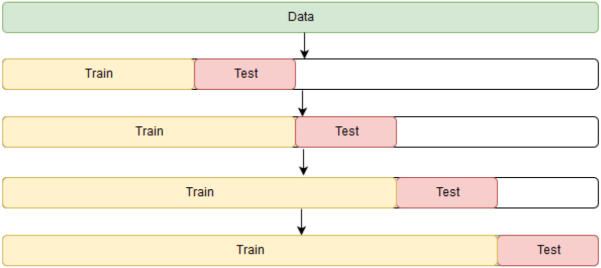
- Universe of assets:
  - S&P 1500
  - 50 largest cap companies in the US
  - Indices / ETFs (*e.g.*, SPDR ETFs)
- Idea 1: Sparse Optimization Methods for Cointegrated baskets
  - Daily tick sizes, sub-selection of stocks with optimization constraints
- Idea 2: Lagged Correlation Graphs for Cointegrated baskets
  - 1min tick sizes, sub-selection of stocks by pruning correlation graph

Criteria

- Max drawdown
- Sharpe ratio
- Overall return
- Rolling portfolio beta

Validation

In order to avoid over-fitting problems, and as we want to take into account the non-stationarity of our data, we develop the following validation scheme to test our model:

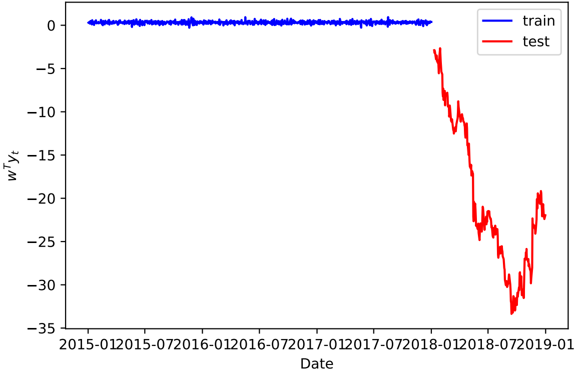


Idea 1: Sparse optimization methods

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{T}(w^T y_t - \mu)^2 \\ \text{subject to} & w \in \mathcal{C} \end{array}$$

- $w \in \mathbf{R}^m$ is the optimization variable
- $\mathcal{C}$ encodes other constraints (*e.g.*, market neutrality, $|\beta^T w| \leq \epsilon$)
- For convex $\mathcal{C}$, this is a convex optimization problem

Problem



- $\mathcal{C} = \mathbf{R}^m$
- Naive method badly overfits (perfect in train, completely unusable in test)

Idea 1: Sparse optimization methods

- Let $w^\star$ be the minimizer of

$$\begin{array}{ll} \text{minimize} & \sum_{t=1}^{T} (w^T y_t - \mu)^2 + \lambda \|w\|_1 \\ \text{subject to} & w \in \mathcal{C} \end{array}$$

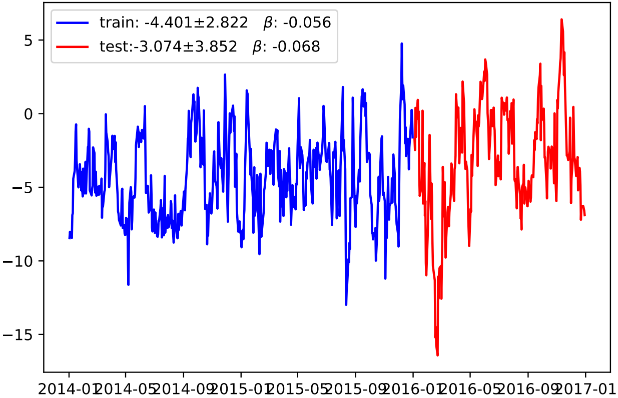
- Incorporate *polishing*, works well in practice

Idea 1: Example

- Energy sector, $m = 28$ stocks (including SPY, XLE)
- Market neutrality constraint
- Train/validate on Jan 2014 - Jan 2016, test on Jan 2016 - Jan 2017.

```
Nonzero weights:
APA 0.39888001844326315
COG -0.4113619169466383
CVX -0.8906814059244456
DVN 0.2599217257736349
EQT -0.37149490868708585
FTI 0.7475351494216929
HAL 0.41189146949001026
HP -0.26619894347392736
KMI -0.7567502256169396
MPC -1.1515298928003135
MRO 1.3938999332485456
NBL 0.6427229074945807
NOV -0.5664895524348179
OXY -0.9299458067924226
PXD -0.1457447179318823
RRC 0.4317833087790374
SLB -0.8402034439359056
SPY 1.0
```

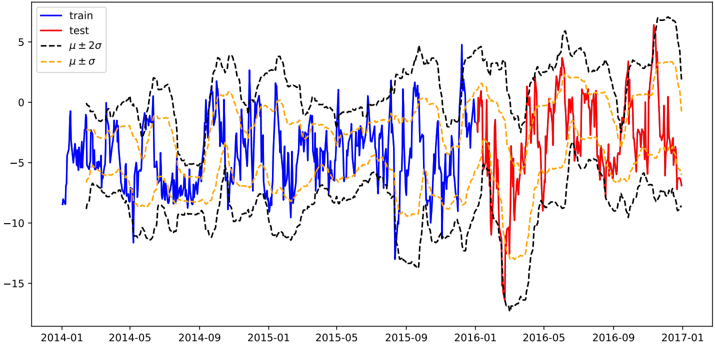
Idea 1: Example



Idea 1: Example

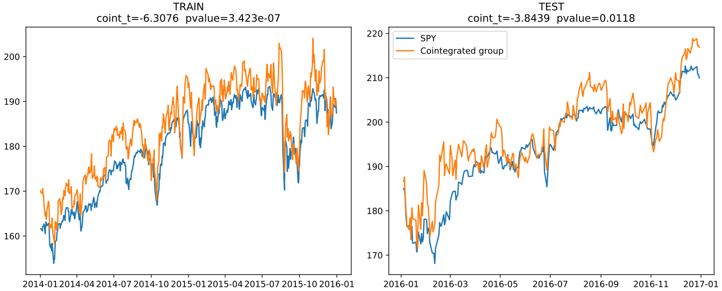
- Can be long 1 share, short 1 share, or hold nothing
- Short/long 1 share of weighted portfolio when above/below inner band
- Run policy until either
  - Get to end of test set
  - $w^T y_t \notin [\mathbf{mean}(w^T y_t) \pm 2 \cdot \mathbf{std}(w^T y_t)]$
    - rolling, 30-day backward
  - Liquidate inventory at end (if needed)

Idea 1: Example



- On 2016
  - 54 trades (enters/exits)
  - $\approx$ 16% return, 8% drawdown

Idea 1: Example

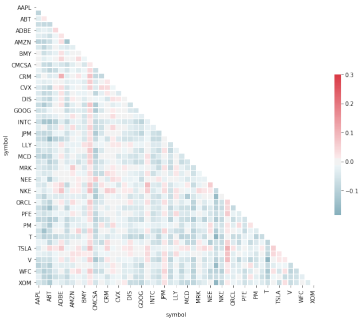


- Just for fun

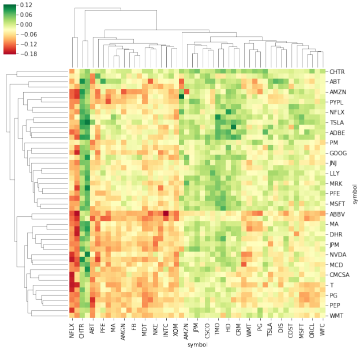
Idea 2: From Correlation to Co-Integration

- Assume returns of stock P_t are correlated with the lagged returns of stock Q_{t+dt} for a given dt.
- Assuming that both P and Q have no alpha, if the return of P is excessively large, we want to short P and go long Q, as Q returns is expect to catch up and P's returns is expected to go down.
$\longrightarrow$ we get a strategy that is similar to the co-integration's one

Idea 2: Lagged correlation



(a) Lagged Correlation Matrix

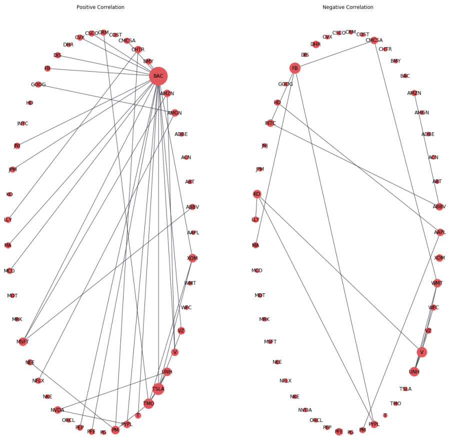


(b) Correlation Clusters

Figure: First Lagged Correlation Analysis for 15-minute data (2016)

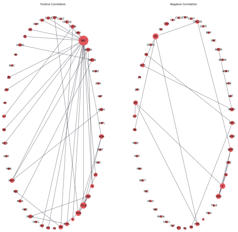
- This look promising for finding correlated baskets

Idea 2: Establishing Stock Universe from Lagged Correlation Graph

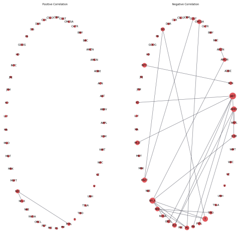


Selection Process:

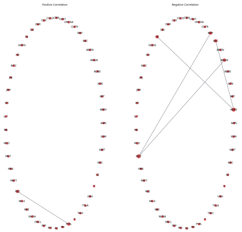
- Shuffle data rows
- Compute the new Lagged Correlation
- Stocks for which the lagged corrrelation is the top/bottom 5% are considered to be significantly correlated
- We apply a Bonferroni correction to account for the multiple tests



(a) 1 minute



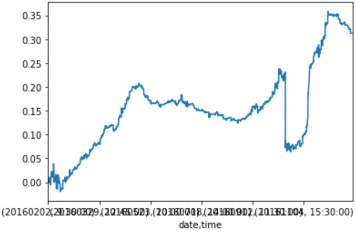
(b) 15 minutes



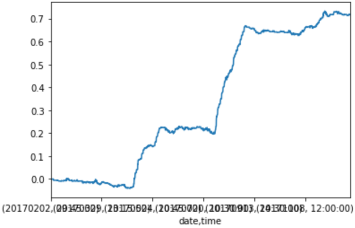
(c) 1 day

Idea 2: Strategy Description

- For each connected component in Lagged Correlation Graphs, create a basket
- For each basket, regress (with Ridge) the most connected node against the other
- The *difference* is assumed to be mean-reverting
- We rebalance our betas every month, and backtest the PnL and the next



(a) 2016-2017



(b) 2017-2018

Figure: Strategy Backtest

Conclusion

- Sparse optimization methods tended to create baskets of stocks that exhibited *tradable* mean-reverting properties
- Lagged correlation graphs were effective at finding correlated stocks
- Two very different ways of finding baskets